Blind source separation of (non-Gaussian) graph signals

Jari Miettinen, Sergiy Vorobyov, Esa Ollila

Aalto University, Finland

ICASSP 2020

Motivation

- There are blind source separation (BSS) methods for non-Gaussian data, time series, matrix- and tensor-valued data, functional data, etc.
- Here we focus on signals on graphs. BSS on graphs has got only a little attention so far. We propose
  - modifications to the one graph BSS method which has been introduced
  - new method which uses non-Gaussianity and dependencies in the data which are given by graphs.

Outline

Blind source separation model

We consider real-valued and linear BSS model for the graph data

$$\boldsymbol{X} = \boldsymbol{\Omega}\boldsymbol{Z},$$

where

- $\boldsymbol{X}$ is the $P \times N$ data matrix,
- $\boldsymbol{\Omega}$ is a full rank unknown $P \times P$ mixing matrix,
- $\boldsymbol{Z}$ is a $P \times N$ matrix of latent independent components, which have zero means and unit variances.

- The aim is to estimate $\mathbf{\Gamma} = \mathbf{\Omega}^{-1}$ using only $\boldsymbol{X}$.

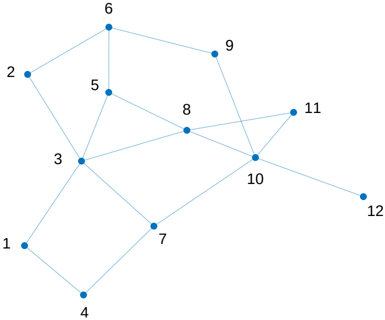
1. Whitening:
$$\tilde{\mathbf{X}} = \hat{\boldsymbol{S}}_0(\boldsymbol{X})^{-1/2}(\boldsymbol{X} - \bar{\boldsymbol{X}})$$
where $\hat{\boldsymbol{S}}_0$ is the sample covariance matrix.

2. Choose how to measure the independence.
3. Find such orthogonal matrix $\hat{\boldsymbol{U}}$ that $\hat{\boldsymbol{U}}\tilde{\mathbf{X}}$ has maximally independent components with respect to the measure.

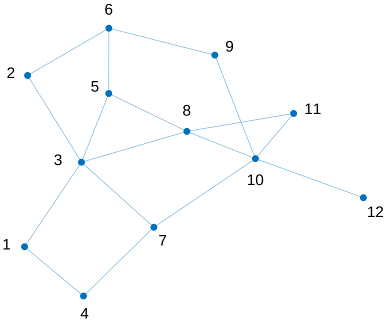
4. The unmixing matrix estimate is $\hat{\mathbf{\Gamma}} = \hat{\boldsymbol{U}}\boldsymbol{S}_0^{-1/2}$.

- Sources can be separated with different strategies:
  - Use marginal distributions of the components. Make them as non-Gaussian as possible.
  - Use correlations of the data points. Time series, matrix, tensor or graph data.
- Here we combine methods which use non-Gaussianity and graphs.

2. Graph signals



Undirected graph with $N = 12$ nodes can be presented by $N \times N$ adjacency matrix $\mathbf{A}$.



$$A = \begin{pmatrix} 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \end{pmatrix}$$

Graph moving average (GMA) model

- GMA signal model of order m, GMA(m), is given by

$$\mathbf{z} = \mathbf{y} + \sum_{l=1}^{m} \theta_l \mathbf{A}^l \mathbf{y}$$

where $\mathbf{y} \triangleq [y_1, \ldots, y_n]^T$ with $y_1, \ldots, y_n \sim N(0, \sigma_y^2)$ and mutually independent, and $\theta_1, \ldots, \theta_m$ are the MA coefficients.

- In GMA(1) model the ith node is given by

$$z_i = y_i + \theta \sum_{j \in \mathcal{N}_i} a_{ij} y_j,$$

where $\mathcal{N}_i$ denotes the incoming neighbors of node i. Thus, if $\theta \neq 0$, z_i and z_j are correlated if ith and jth nodes are neighbors, or if they have common incoming neighbors.

3. BSS of graph signals

Graph autocovariance matrix

- Let us define the sample graph autocovariance matrix at lag k with respect to adjacency matrix $\mathbf{W}$ as

$$\mathbf{S}_k(\mathbf{X}, \mathbf{W}) = \frac{\mathbf{X}\mathbf{W}^k\mathbf{X}^\top}{N}$$

- It satisfies

$$\mathbf{S}_k(\mathbf{U}\mathbf{X}, \mathbf{W}) = \mathbf{U}\mathbf{S}_k(\mathbf{X}, \mathbf{W})\mathbf{U}^\top$$

GraDe (graph decorrelation), Blöchl et al. (2010)

1. Whiten the data $\mathbf{X} \to \tilde{\mathbf{X}}$.
2. Choose symmetric $\mathbf{W}$ and $K \geq 1$.
3. Find orthogonal $\hat{\mathbf{U}}$ that maximizes

$$\sum_{k=1}^{K} \|\mathrm{diag}(\mathbf{U}\hat{\mathbf{S}}_k(\tilde{\mathbf{X}}, \mathbf{W})\mathbf{U}^T)\|^2.$$

4. The GraDe unmixing matrix estimate is $\hat{\Gamma}_K(\boldsymbol{W}) = \hat{\boldsymbol{U}}\boldsymbol{S}_0^{-1/2}$.

We modify GraDe in two ways:

1. Allow multiple adjacency matrices $\mathbf{W}_1, \ldots, \mathbf{W}_M$.
2. Standardize the graph autocovariance matrices so that the scales of $\mathbf{W}_1, \ldots, \mathbf{W}_M$ do not have any effect:

$$\tilde{\mathbf{S}}_k(\mathbf{X}, \mathbf{W}_m) = \frac{\mathbf{X}\mathbf{W}_m^k\mathbf{X}^\top}{\|\mathbf{W}_m^k\mathbf{X}^\top\|}$$

4. Graph JADE

Graph JADE

- Objective function of Graph JADE to be maximized is

$$f_{\mathrm{GraphJADE}} = b \sum_{m=1}^{M} \sum_{k=1}^{K} \|\mathrm{diag}(\mathbf{U}\tilde{\mathbf{S}}(\tilde{\mathbf{X}}, \mathbf{W}_m^k)\mathbf{U}^\top)\|^2$$
$$+ (1-b) \sum_{k=1}^{P} \sum_{l=1}^{P} \|\mathrm{diag}(\mathbf{U}\hat{\mathbf{C}}^{k,l}\mathbf{U}^\top)\|^2$$

  where $b \in [0,1]$ is a weight parameter and

$$\hat{\mathbf{C}}^{k,l} = \sum_{i=1}^{N} [\tilde{\mathbf{X}}]_{k,i} [\tilde{\mathbf{X}}]_{l,i} [\tilde{\mathbf{X}}]_{\cdot,i} [\tilde{\mathbf{X}}]_{\cdot,i}^\top - \mathbf{E}_{P\times P}^{k,l} - \mathbf{E}_{P\times P}^{l,k} - \mathrm{tr}\{\mathbf{E}_{P\times P}^{k,l}\}.$$

- JADE (Joint Approximate Diagonalization of Eidenmatrices) (Cardoso and Soloumiac, 1993) is based on joint diagonalization of fourth-order cumulant matrices. Therefore, implementation of Graph JADE is simple.

Identifiability conditions for Graph JADE

- Let $\mathcal{W}$ denote the set of adjacency matrices in Graph JADE. Then the identifiability conditions can be stated as follows.
- For any pair of ICs $\boldsymbol{z}_i$ and $\boldsymbol{z}_j$
  (i) there is a matrix $\mathbf{W} \in \mathcal{W}$ such that $\mathbb{E}\{\text{diag}(\tilde{\mathbf{S}}(\mathbf{Z}, \mathbf{W}))_i\} \neq \mathbb{E}\{\text{diag}(\tilde{\mathbf{S}}(\mathbf{Z}, \mathbf{W}))_j\}$

  or

  (ii) $\mathbb{E}\{N^{-1} \sum_{n=1}^{N} z_{in}^4\} \neq 3$ or $\mathbb{E}\{N^{-1} \sum_{n=1}^{N} z_{jn}^4\} \neq 3$
- Notice that (i) alone gives the identifiability conditions for GraDe and (ii) alone for JADE.

5. Simulations

Performance measure in simulation studies

- Minimum distance (MD) index (Ilmonen et al., 2010b)

$$D(\hat{\mathbf{\Gamma}}) = \frac{1}{\sqrt{p-1}} \inf_{\mathbf{C} \in \mathcal{C}} \|\mathbf{C}\hat{\mathbf{\Gamma}}\mathbf{\Omega} - \mathbf{I}_p\|$$

  where

  $\mathcal{C} = \{\mathbf{C} \ : \ \text{each row and column of } \mathbf{C} \text{ has exactly one non-zero element}\}.$

- $0 \leq D(\hat{\mathbf{\Gamma}}) \leq 1$, smaller value is better.
- $N(P-1)D(\hat{\mathbf{\Gamma}})^2$ will be used as the final measure.

Simulations

- Four estimators are compared in four models:
  - GraDe (modified), $K = 2$
  - JADE
  - Symmetric FastICA (Hyvärinen, 1999), *tanh nonlinearity function*
  - *Graph JADE*, $K = 2$, $b = 0.8$
- The figures will show averages of 1000 runs.

Model 1

- Four independent components are GMA(1) signals with joint $\boldsymbol{A}$ from Erdös–Rényi model $G(N, 0.05)$, and

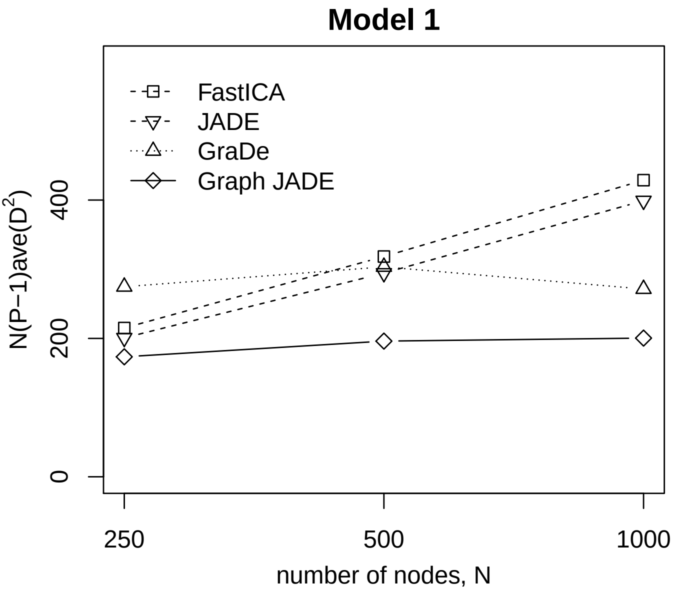
  IC1 $\theta = 0.02$, y follows t_5-distribution

  IC2 $\theta = 0.04$, y follows t_{10}-distribution

  IC3 $\theta = 0.06$, y follows t_{15}-distribution

  IC4 $\theta = 0.08$, y follows Gaussian distribution

- In this model, sources can be separated using non-Gaussianity or graph decorrelation.


Model 1
FastICA
JADE
GraDe
Graph JADE
N(P–1)ave(D²)
0
200
400
250
500
1000
number of nodes, N

Model 2

- Four independent components are GMA(1) signals with joint $\boldsymbol{A}$ from Erdös–Rényi model $G(N, 0.05)$,

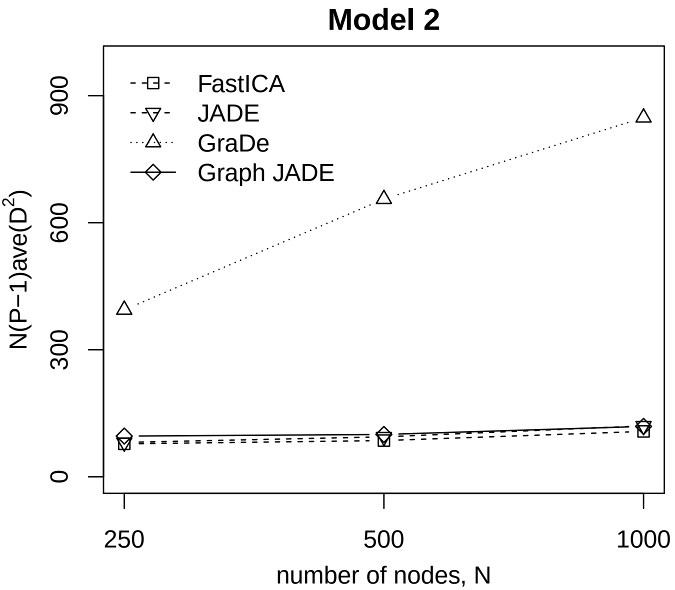
  IC1 $\theta = 0.05$, y follows t_5-distribution

  IC2 $\theta = 0.06$, y follows uniform distribution

  IC3 $\theta = 0.07$, y follows exponential distribution with $\lambda = 1$

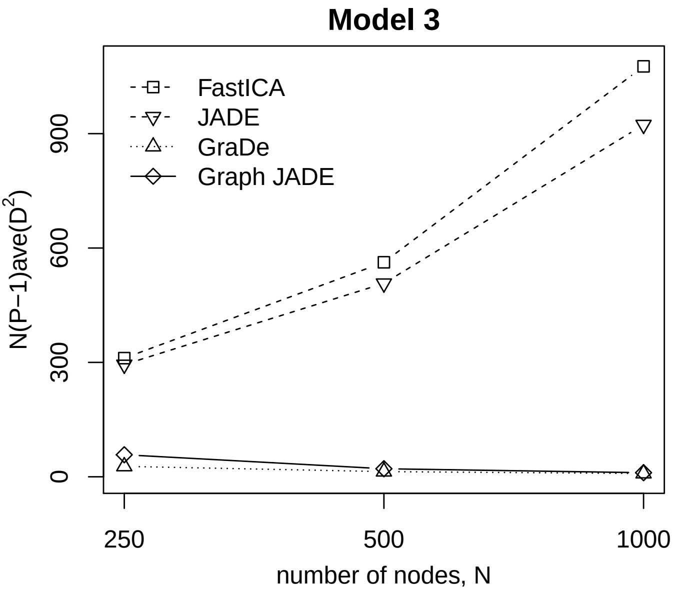
  IC4 $\theta = 0.08$, y follows Gaussian distribution

- This model is challenging for GraDe.


Model 2
FastICA
JADE
GraDe
Graph JADE
0
300
600
900
N(P−1)ave(D^2)
250
500
1000
number of nodes, N

Model 3

- Four independent components are GMA(1) signals with different adjacency matrices. For all components $\theta = 0.05$ and distribution of y is t_{15}.
- This model is challenging for FastICA and JADE.



Model 3
FastICA
JADE
GraDe
Graph JADE
0
300
600
900
N(P−1)ave(D2)
250
500
1000
number of nodes, N

Model 4

- Four independent components are GMA(1) signals with joint $\boldsymbol{A}$ from Erdös–Rényi model $G(N, 0.05)$, and

  IC1 $\theta = 0.05$, y follows t_5-distribution

  IC2 $\theta = 0.05$, y follows Gaussian distribution

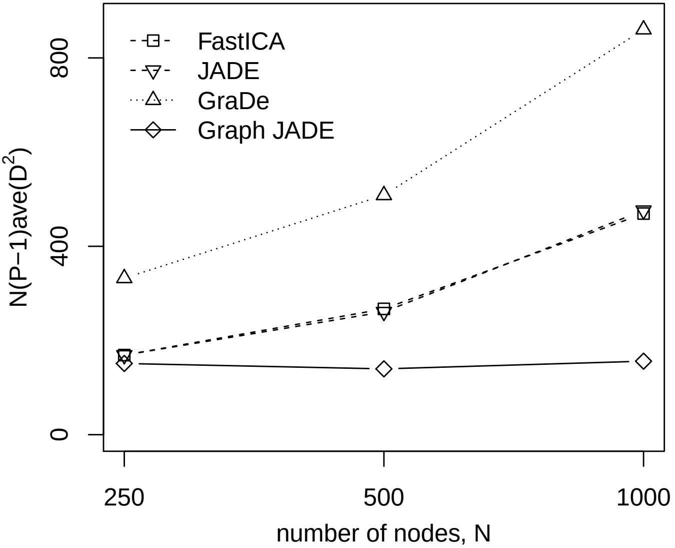
  IC3 $\theta = 0.1$, y follows uniform distribution

  IC4 $\theta = 0.1$, y follows Gaussian distribution

  $\theta = 0.05, 0.05, 0.1, 0.1$ and distributions of ys are t_5, Gaussian, uniform, and Gaussian, respectively.

- In this model, there are two components (IC2 and IC4) which can not be separated using FastICA or JADE, and two pairs of components (IC1 and IC2, IC3 and IC4) which can not be separated using GraDe.

Model 4


FastICA
JADE
GraDe
Graph JADE
800
400
0
N(P−1)ave(D2)
250
500
1000
number of nodes, N

Conclusions

- We have proposed a BSS method called Graph JADE for graph signals, which use both possible non-Gaussianity and graph dependencies within the latent components.
- The proposed method outperforms JADE, FastICA, and GraDe when the sources are non-Gaussian graph signals, and meets the performance of the best method with minor losses when the sources exhibit only non-Gaussianity or graph dependence.

References

1. Blöchl, F., Kowarsch, A. and Theis, F. J. (2010). Second-order source separation based on prior knowledge realized in a graph model, in "Latent Variable Analysis and Signal Separation", LNCS (V. Vigneron, V. Zarzoso, E. Moreau, R. Gribonval, and E. Vincent, eds.), vol. 6365, 434–441, Springer, Heidelberg.

2. Cardoso, J. and Soloumiac, A. (1993) Blind beamforming for non-gaussian signals, *IEE Proceedings-F*, vol. 140, 362–370.

3. Hyvärinen, A. (1999) Fast and robust fixed-point algorithms for independent component analysis, *IEEE Transactions on Neural Networks*, vol. 10, no. 3, 626–634.

Thank you!